A biologically inspired immunization strategy for network epidemiology

Yang Liu[a], Yong Deng[a,b,*], Marko Jusup[c,*], Zhen Wang[d,e,*]

[a]*School of Computer and Information Science, Southwest University, Chongqing 400715, China*
[b]*School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China*
[c]*Center of Mathematics for Social Creativity, Hokkaido University, Sapporo 060-0812, Japan*
[d]*School of Automation, Northwestern Polytechnical University, Xi'an 710072, China*
[e]*Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Fukuoka, 816-8580, Japan*

Abstract

Well-known immunization strategies, based on degree centrality, betweenness centrality, or closeness centrality, either neglect the structural significance of a node or require global information about the network. We propose a biologically inspired immunization strategy that circumvents both of these problems by considering the number of links of a focal node and the way the neighbors are connected among themselves. The strategy thus measures the dependence of the neighbors on the focal node, identifying the ability of this node to spread the disease. Nodes with the highest ability in the network are the first to be immunized. To test the performance of our method, we conduct numerical simulations on several computer-generated and empirical networks, using the susceptible-infected-recovered (SIR) model. The results show that the proposed strategy largely outperforms the existing well-known strategies.

Keywords: infectious agent, *Physarum polycephalum*, heterogeneous topology, degree centrality, betweenness centrality, closeness centrality, SIR model

1. Introduction

Infectious agents, in the broadest sense, can spread in a 'population' of humans, animals, and nowadays technological devices. In the case of an outbreak, the goal is to minimize the damage to the 'population' without violating the constraints of a limited budget. Achieving this goal in 'populations' other than the primitive, well-mixed ones is the main reason why the problems of (i) epidemics spreading in complex networks [1, 2, 3, 4, 5] and (ii) the corresponding immunization strategies [6, 7, 8, 9] attracted considerable attention in the literature. Such a trend is further spurred by the outcomes of ineffectively controlled epidemics—exemplified by the severe acute respiratory syndrome (SARS) and the swine flu [8]—which rapidly spread all over the world due to globalization and better means of transport [10, 11]. An immediate concern, therefore, is the development of the effective countermeasures against epidemics, wherein immunization science plays an important, if not critical, role.

Multiple immunization strategies have been considered for complex networks. *Uniform* or *random* immunization [12] selects any node in the network with equal probability at each time step, which results in a strategy suitable for homogeneous networks, but ignores the fact that in highly heterogeneous networks eradicating an infective agent cannot be guaranteed regardless of the fraction of immunized nodes. *Targeted* immunization [13, 14] seeks to overcome this problem by selecting the nodes with the highest ability to spread the disease. Such ability is often measured in terms of degree centrality, defined as the number of ties a node has, or betweenness centrality, which indicates how often a node is located along the shortest route between two other nodes. A strategy based on degree centrality is highly effective, yet it neglects the structural significance of a node because the most connected nodes are not necessarily the ones that facilitate disease propagation from one dense cluster to another [9]. Betweenness centrality can be used to overcome such a shortcoming, but suffers from a considerable computational cost [9]. In general, targeted immunization requires global knowledge of network properties, which is impractical. Motivated by such an impracticality, so-called *acquaintance* immunization [15, 16] was developed, whereby immunized nodes comprise only a small fraction of random neighbors of randomly selected nodes. Furthermore, a number of alternative and/or more specialized strategies have also

*Corresponding authors
Email addresses: professordeng@163.com (Yong Deng), mjusup@gmail.com (Marko Jusup), zhenwang0@gmail.com (Zhen Wang)

been discussed [17, 18, 19, 20, 21]. Owing to their immediate practical and economic implications, additional studies involving novel immunization strategies are still of considerable importance.

Herein we propose an efficient biologically inspired immunization strategy for network epidemiology. The term *biologically inspired* is used in the sense that we draw inspiration from a single-celled, amoeba-like organism *Physarum polycephalum*, capable of transporting signals and nutrients through a dendritic (i.e. tree-like) network of tubular structures called pseudopodia [22]. *Physarum* is particularly interesting because of the ability to perform cellular computations of a sort, which lead to the solution of, for example, the shortest path problem. This and similar observations served as a basis for constructing a mathematical model of an adaptive transport network driven by Poiseuille flow that exibits computational abilities much like *Physarum* itself [23]. Subsequently, a *Physarum*-type algorithm has been used for solving the Steiner Minimum Tree problem [24, 25], for optimizing network design [26, 27, 28, 29], and for a variety of other applications [30, 31, 32, 33, 34].

The paper is organized as follows. In Section 2 we describe how to construct a *Physarum* model and modify it into a strategy for immunizing complex networks. Model properties and performance are examined in Section 3. In particular, we emphasize (i) the difference in functioning between the proposed strategy and other, well-known counterparts, (ii) the gain in performance by accounting for the structural importance of nodes, and (iii) the ability to capture the key attributes of a whole network based solely on the successive use of local information. Finally, in Section 4 we summarize the take-home message and discuss the potential drawbacks, as well as the future developments.

2. Methods

The proposed, biologically inspired strategy (denoted simply AS, where A stands for amoeba) is constructed in three steps. First, we describe the *Physarum* model for the shortest path selection. We then modify the model by adding a noise factor to select the paths alternative to the shortest one along which the disease can still spread, albeit with a lower probability. Second, the original *Physarum* (single-source, single-sink) model is further modified to become a single-source, multi-sink model to consider the spread of the disease to multiple targets simultaneously. With these modifications, the model measures the dependence of each node in the network on the focal node. The third and final step in the construction of AS is motivated by the fact that applying the modified *Physarum* model to a large network is impractical due to a possible lack of information on the network structure or the high computational expense. Accordingly, for each focal node, a subnetwork consisting of the focal node itself and its R-step neighbors is separated from whole network, whereupon the model is applied to this R-local subnetwork. To test the performance of AS, we conduct numerical simulations on several computer-generated and empirical networks.

2.1. A Physarum *model for the shortest paths selection*

The shortest path selection by *Physarum polycephalum* is based on the transformations of tubular structures (pseudopodia) [23, 35] and a positive feedback from flow rates [36, 37]. Namely, high rates of the protoplasmic flow stimulate the diameter of pseudopodia to increase, whereas at low flow rates the diameter tends to decrease. The pseudopodia thickness thus adapts to the flow rate. More formally, we observe a set of nodes N, containing two end-nodes, i_1 and i_2 (for biological reasons also called food-source nodes), and any number of in-between nodes, $i \in N$. The edge connecting nodes $i, j \in N$ is denoted as i-j and serves as a symbolic representation of a pseudopodium. Denoting by Q_{ij} the flux through edge i-j (i.e. from i to j) and assuming the flow through pseudopodia to be approximately Poiseuille flow, we obtain:

$$Q_{ij} = \frac{\pi r_{ij}^4}{8\xi L_{ij}}(p_i - p_j) = \frac{D_{ij}}{L_{ij}}(p_i - p_j), \tag{1}$$

where ξ is the viscosity coefficient of the sol, $D_{ij} = \pi r_{ij}^4/8\xi$ is a measure of the conductivity of pseudopodia (with r_{ij} being the radius), and p_i is the pressure at node i. L_{ij} is the length of edge i-j. One interpretation for these lengths follows from the fact that L_{ij}^{-1} determines the importance of edge i-j, which means that in an unweighted network $L_{ij} = 1$ for all possible is and js.

Pressures, p_i in Eq. (1) are unknown. To obtain them, we further assume that in-between nodes are of zero capacity and that the sol conservation law is upheld, yielding at each node j, $\sum_{i\in\Gamma(j)} Q_{ij} = 0$, $j \neq 1, 2$, where i runs across the set of the nearest neighbors of j, $\Gamma(j)$. For the source node, 1, and the sink node, 2, we have $\sum_{i\in\Gamma(1)} Q_{i1} = -I_0$ and $\sum_{i\in\Gamma(2)} Q_{i2} = I_0$, respectively, where I_0 is the influx from the source node (or into the sink node). It follows:

$$\sum_{i\in\Gamma(j)} \frac{D_{ij}}{L_{ij}}(p_i - p_j) = \begin{cases} -1 & \text{for } j = 1, \\ +1 & \text{for } j = 2, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where without any loss of generality $I_0 = 1$. By further setting $p_2 = 0$ as the basic pressure level, all p_i's can be determined from the above system of equations. Finally, using Eq. (2) all flows, $Q_{ij} = D_{ij}(p_i - p_j)/L_{ij}$, can also be obtained.

Experiments show that pseudopodia with higher flux rates are reinforced, while those with lower flux rates degenerate [23, 35]. To accommodate this adaptive behavior of pseudopodia, the conductivities, D_{ij}, are assumed to change according to the following equation:

$$\frac{d}{dt}D_{ij} = f(|Q_{ij}|) - aD_{ij}, \tag{3}$$

where a is the decay rate. The functional form $f(|Q_{ij}|)$ is, for simplicity, set to $f(|Q_{ij}|) = b|Q_{ij}|$, where b is a constant [23]. We further simplify the notation by choosing the units in which $b/a = 1$, implying that flow and conductivity have the same dimensions, that pressure has the dimension of length, and that the time, $t' = at$, is dimensionless (in the following, we will suppress the prime because no confusion can occur). To solve Eq. (3), a semi-implicit scheme is used as follows:

$$\frac{D_{ij}^{n+1} - D_{ij}^{n}}{\delta t} = |Q_{ij}^{n}| - D_{ij}^{n+1}, \tag{4}$$

where δt is a time step and the upper index, n, indicates the current moment in time. In this way, the time-varying conductivity, D_{ij}, degenerates to zero for every path between the end-nodes, except the shortest one.

2.2. A noise factor for irrational path selection

The described *Physarum* model solves the shortest path problem, but that is not where our interest lies. To formulate an immunization strategy, it is necessary to consider the paths other than the shortest one. However, in the original *Physarum* model, the flow Q_{ij} from node i to node j tends to unity if the edge i-j is lying on the shortest path or tends to zero otherwise [38]. To account for other pathways as well, we rewrite Eq. (4) as follows:

$$\frac{D_{ij}^{n+1} - D_{ij}^{n}}{\delta t} = \frac{|Q_{ij}^{n}|}{1 - \gamma + \gamma|Q_{ij}^{n}|} - D_{ij}^{n+1}, \ (0 < \gamma \leq 1) \tag{5}$$

where γ is a noise factor. Similar to the noise in game theory [39, 40], γ is introduced to disturb the mass flow distribution through the network's edges. Using the terminology from evolutionary biology, γ permits irrational selection. The modified model approaches the original as $\gamma \to 0$.

2.3. Multi-foraging

For the purpose of devising an immunization strategy, in addition to modifying the model by a noise term, we wish to simultaneously consider the potential for the spread of the disease from a source node to multiple other nodes in the network. To achieve such a feat, we rewrite Eq. (2) as

$$\sum_{i \in \Gamma(j)} \frac{D_{ij}}{L_{ij}}(p_i - p_j) = \begin{cases} -1 & \text{for } j = 1, \\ \frac{1}{n-1} & \text{for } j = \{2, 3, \cdots, n\}, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where n is the number of nodes in the network. We call this model a multi-forage model because multiple sink-nodes can "forage" on a single source-node. By applying the multi-forage model, we are in a position to derive a proxy for the ability of the focal node to spread the disease.

2.4. A biologically inspired immunization strategy (AS)

The extended *Physarum* model has all the desired properties for formulating an immunization strategy, yet the information on the structure of the whole network may not exist. Even if such information is available, the model may not be applicable to a very large network for computational reasons [41]. We therefore pursue an R-local approach that considers only a subnetwork of the focal node comprised of the focal node's R-level neighbors and all the accompanying edges. Specifically, a proxy for the focal node's ability to spread the disease is obtained through the following steps:

1. Take $\Gamma_R(k)$, the set of all R-level neighbors of the focal node k.

2. For each $l \in \Gamma_R(k)$, one iteration of the model is run in which node l is the source and all other $j \in \Gamma_R(k) \cup \{k\}$ are the sinks. Each iteration thus generates inflows (i.e. positive flows) $Q_{ij}^{(k,l)}$ along all edges i-j in the selected R-local subnetwork of focal node k.

3. The summation across all iterations, $Q_{ij}^{(k)} = \sum_{l \in \Gamma_R(k)} Q_{ij}^{(k,l)}$, gives the so-called flow matrix.

4. Finally, the ability of focal node k to spread the disease is quantified by a score:

$$AS(k) = \sum_{i\text{-}j} Q_{ij}^{(k)}, \tag{7}$$

where i-j symbolizes the summation over all the edges in the R-local subnetwork of focal node k.

5. The focal node with the highest AS-score is the first to be immunized.

The first three steps of the described strategy may be somewhat difficult to understand using just a mathematical notation. Accordingly, these steps are illustrated in Fig. 1. To distinguish focal node 1 from its first-level neighbors, we mark it with a wax-yellow color. The results of a single model iteration, in which node 3 is chosen as the source node, are shown in Fig. 1a. This iteration is immediately followed by another one (Fig. 1b), such that node 2 is the source node. The results are summed to produce the flow matrix in Fig. 1c. The AS-score for focal node 1 is obtained by repeating the process for all of 1's first-level neighbors and then summing the elements of the flow matrix across all edges.

3. Results and discussion

Next, we examine several interesting properties of AS and test its performance against strategies based on degree centrality (DCS), betweenness centrality (BCS), and closeness centrality (CCS). Both stylized, computer-generated networks and real-world, empirical networks are used in the performance tests. Additionally, the success of immunization is tested in conjunction with a susceptible-infectious-recovered (SIR) model.

3.1. Properties of AS

The role of the noise factor and irrational path selection. The proposed immunization strategy is designed to quantify the ability of a focal node to spread the disease through the network. Put alternatively, we want to know (i) the extent to which the neighbors depend on the focal node and (ii) the relative importance of disease spreading pathways on which the focal node lies. Nodes 2, 3, and 4 in the inset of Fig. 2, for example, are just partially dependent on node 1 because they are, among others, connected by pathways excluding the latter node. Node 5, by contrast, is fully dependent on node 1. Furthermore, if we consider the spread of the disease from node 5 to node 3, path 5-1-3 is the most direct, but the disease can spread through alternative paths 5-1-2-3 and 5-1-4-3. Such alternatives need to be taken into account to properly determine the ability of the focal node to spread the disease. The original *Physarum* model cannot yield the desired result. If we imagine for a moment that the link between nodes 1 and 3 is removed and assume that path 5-1-2-3 is just a tiny bit shorter than alternative path 5-1-4-3, all flow would be assigned to the former path. Node 4 would consequently be disregarded as irrelevant for the spread of the disease, while nodes 1 and 2 would (wrongly) be considered of equal relevance. However, the possibility of having an epidemic spread via both paths (i.e. 5-1-2-3 and 5-1-4-3) would suggest that nodes 2 and 4 should be seen almost as equivalents, whereas node 1 is of more immediate concern. The role of the noise factor, γ, is to resolve the described problem. Indeed, the three paths from node 5 to node 3 are properly distinguished in Fig. 2 based on the flows, if the value of the noise factor is set to some $\gamma > 0.5$.

The advantage of AS over the other comparable strategies. To exemplify why AS should perform well at immunizing a networked population, Fig. 3 shows three networks in which the wax-yellow circle represents the focal node, whereas the other circles are its 1-step neighbors. The degree centrality strategy (DCS) fails to distinguish nodes 1 and i in Figs. 3a and b because it neglects how neighbors are connected among themselves. The betweenness centrality strategy (BCS) solves this problem, but by considering only the shortest paths, the strategy fails to distinguish nodes i and a in Figs. 3b and c. Using AS, however, we obtain $AS(1) = 5.4$, $AS(i) = 8.0$, and $AS(a) = 9.2$, respectively, which highlights the capacity of the proposed strategy to classify nodes for immunization where the other comparable strategies leave us in the dark.

To further emphasize how different AS is from the other well-known strategies, we compare the results obtained using a simple, yet non-trivial network taken from Ref. [42]. Fig. 4 shows the first five nodes marked for immunization by each of the strategies. DCS and CCS, in fact, end up marking the same five nodes for immunization shown in Fig. 4a, albeit in a different order. Note, however, that DCS is ambiguous as to what the order should be because nodes 3 to 6 have the same degree. Without an additional criteria, there is no reason to prefer one node over the other (we actually followed the numbering of nodes). Although all strategies singled out nodes 1 and 16, this is where the similarities end. BCS and AS (Figs. 4b and c) go beyond the node

degree to mark nodes 23 and 24, both with only two links, for immunization. Nonetheless, the best performer is AS because after the first five nodes are immunized, the biggest cluster that can get infected is of size three (containing nodes 2, 3, and 4), whereas the other strategies leave clusters of size four (DCS and CCS) and even size five (BCS) exposed.

3.2. Performance testing in stylized, computer-generated networks

To test the performance of AS, we plot the fraction of a network that can be infected, F, against the fraction of immunized nodes, q. The tests are first performed in four stylized, computer-generated networks, the two of which are of Erdős-Rényi (i.e. random) type, whereas the remaining two are of Albert-Barabási (i.e. scale-free) type [43, 44]. The results are shown in Fig. 5. Although AS performs very well, its performance is matched by DCS, which may seem surprising at first. However, random networks are not supposed to have structurally important nodes in the sense of Ref. [9] (i.e. nodes that connect dense clusters, but do not necessarily have a high degree themselves). In scale-free networks, by contrast, structurally important nodes are most likely the ones with a high degree. These results suggest that random and scale-free networks may be overly stylized for thorough performance testing of the considered immunization strategies.

3.3. Performance testing in real-world, empirical networks

For the purpose of additional performance testing, we rely on four empirical networks as follows:

- *Email*, a complex email network of an organization with over 1000 employees [45];
- *GR-QC*, a collaborative scientific network between more than 5000 authors who submitted their articles to General Relativity and Quantum Cosmology category of arXiv e-print archive [46].
- *Power grid*, a representation of connections between almost 5000 power stations in the western US [47];
- *Yeast*, a protein-protein interaction network of yeast containing over 2000 proteins [48].

Aside from the fact that all these networks are well-documented and readily obtainable, the first two were selected for their relevance to the spread of computer viruses and the possibility of sharing some properties with the networks of human physical contacts. Power grid network is interesting in the context of the protection from cascading failures, whereby 'immunized' nodes should have redundant capacity to handle the extra load when another network element fails. 'Immunizing' Yeast network is admittedly an academic exercise, yet a useful one because the properties of empirical networks are not readily observed and analyzed in their stylized, computer-generated substitutes [44]. The descriptive statistics of the four selected networks is summarized in Table 1.

The results of performance testing in real-world, empirical networks are shown in Fig. 6. AS is seen outperforming all three well-known strategies to which it was compared. Moreover, in Figs. 6b-d the performance of AS diverges from that of DCS, thus confirming our previous comment that random and scale-free networks are overly stylized to reveal the key properties of the considered immunization strategies necessary for a thorough evaluation. Another important point is that AS was designed to combine the qualities of DCS and BCS, meaning that AS should perform well when DCS is better than BCS (which is precisely the case in Fig. 5, for example), but also when it is the other way around (as exemplified in Figs. 6b and d). The ability of AS to perform as expected gives us some confidence that our design was successful and that AS should outperform the other strategies in an even more dramatic fashion if simulations incorporated the realistic elements of an epidemic.

3.4. Performance testing in conjunction with the susceptible-infectious-recovered model

To incorporate the realistic elements of an epidemic into performance testing, we resort to a susceptible-infectious-recovered (SIR) epidemiological model [49, 50, 51, 52] defined as follows:

- Each node of the network can be in one of the three possible states: susceptible, infected, or recovered;
- Initially, nodes to be immunized are determined and removed from the network, including the incident links;
- A node from which the rest of the network can be infected is randomly selected, whereas all the other nodes start in the susceptible state;
- At each time step, infected nodes transmit the disease to their susceptible neighbors with probability λ;

- Infected nodes recover with probability η, whereupon they get removed from the network.

The last two steps of the process outlined above are repeated until no infected nodes remain.

The results of performance testing in conjunction with the SIR model, where $\lambda = 0.2$ and $\eta = 0.05$, are shown in Fig. 7. The performance is measured in terms of the infected fraction, P_i, and the recovered fraction, P_r, of nodes as the functions of time. The former measure is instantaneous in nature, whereas the latter quantifies the cumulative effect of an epidemic. In a scale-free network, consistent with the results in Fig. 5, DCS is an effective immunization strategy, yet AS performs considerably better (Fig. 7a). In GR-QC network, based on Fig. 6b, one might expect that BCS rather than DCS performs well. Although this is precisely the case, AS achieves an even higher effectiveness (Fig. 7b). In the power network, judging from Fig. 6c, DCS is surprisingly weak, but such a lackluster performance may have been caused by a low fraction of immunized nodes (5%) in the presented example (Fig. 7c). More importantly, AS once again emerges as a clear victor. The same is even more pronounced in the Yeast network (Fig. 7d). In summary, the maximum infected fraction and the final recovered fraction are considerably (2 to 20 times) lower if AS is used instead of DCS, BCS, or CCS with the same fraction of immunized nodes.

4. Conclusions and outlook

Achievements. We designed and tested a new, efficient network immunization strategy by drawing inspiration from an amoeba-like organism, *Physarum polycephalum*. By looking at the dependence between any two nodes, we modified the original *Physarum* model with a noise factor for similar path finding. We then extended the model to include multi-foraging in order to obtain the dependence of other nodes on the focal one. This dependence is expressed in the form of a flow matrix, which indicates the ability of the focal node to spread the disease. Perhaps the most important properties of such a design are the ability to (i) distinguish situations that other comparable methods cannot and (ii) recognize structurally critical nodes that may not have a high degree themselves, yet they connect dense clusters of nodes.

Applicability. In the proposed strategy, immunized nodes are those with a higher ability to spread the disease as indicated by the *AS*-score. However, calculating such a score in reality may be difficult. The information on human physical contacts is hard to come by and even if the necessary information is available, a strong time-dependence is likely to cause erroneous assessments. The proposed method partly overcomes these difficulties by being local in the sense of Ref. [9]. The information on a few R-local subnetworks can be used to start an immunization program, which can later be expanded in scope as the new information comes in. The results are encouraging in this context because, even if AS is applied locally, the attributes of the whole network are successfully revealed.

In view of the difficulties in applying the proposed (and any other) immunization strategy, it is important to admit that more success may be achieved in technological networks for which the relevant, structural information is more readily available. Another context in which the proposed strategy may be more applicable is the spread of a disease in a network of villages or towns, whereby an efficient immunization strategy may identify hotspots for implementing preventive measures with the most impact.

Outlook. Because the effectiveness of immunization programs may be hampered by a lack of information, perhaps the most immediate concern is to design the strategies that (i) perform well under information restrictions and (ii) can combine information from multiple sources. As for the restricted information, AS already functions locally, but achieves efficiency globally. To address the problem of combining several information sources, we may need to resort to the framework of multilayer networks [53], such that each network layer represents one type of information. Upon obtaining the *AS*-score in each layer separately, the results could be combined in a certain fashion to produce one overall score for the decision-making. In fact, optimizing the way in which the results are combined could be an interesting task for the field of intelligent computation [54]. Finally, the incentives for the different behavioral responses to an epidemic could be systematically incorporated into our framework by means of evolutionary game theory [55, 56].

Appendix A: Additional analyses

There are fundamental obstacles to analytically comparing the performance of various immunization strategies. Namely, a network's (mean–field) statistics translate into the statistics of nodes selected for immunization by each of the methods, which in turn determines the value of a common performance indicator (e.g. the largest fraction of nodes that can get infected after immunization, F). This sequence of dependencies is most easily seen by considering the workings of the degree centrality strategy, and is reflected in the fact that the simulated performance of each strategy varies greatly with the network structure. A more important consequence is that any analytical proof that one method is better than the others would have to start by assuming

a mean–field statistic, and therefore would remain specific to such a statistic, largely diminishing the usefulness of this approach. Furthermore, mean–field statistics are generally known only for synthetic networks arising from the well-known (e.g. Erdős–Rényi and scale–free) algorithms, yet these networks fail to fully reveal how advantageous the proposed strategy (AS) is (Fig. 5). AS truly shines when applied to highly heterogeneous, real-world networks (Figs. 6 and 7; also see below).

To strengthen the case for AS, we performed a number of analyses beside those found in the main text. First we compare how AS fares against a fairly recent immunization strategy called k-core (hereafter KCS) [42, 57]. A conceptual advantage of AS in comparison with KCS is that the latter strategy needs global knowledge to decompose a network and then identify the most influential nodes. AS, by contrast, relies only on local identification, meaning that the influence of each node is assessed from the information on the node's immediate neighborhood. Despite the information "handicap" of being local, AS manages to outperform KCS in all four real-world networks considered in the main text (Fig. A1). In addition, we provide the simulation results for the leader rank strategy (LRS) [18] which also appeared in the literature fairly recently. LRS in some instances (Figs. A2a and c) achieves almost the same efficiency as AS, yet we have not identified any examples in which the former strategy would outperform the latter.

To further solidify the leading position of AS, we acquired the information on three additional real-world networks and tested the performance of AS against the rival strategies. Networks used in these tests are:

- *PGP*, a network of mutual trust with more than 10500 involved parties for signing of digital documents using the Pretty Good Privacy algorithm [58, 59, 60];
- *Ca-HepPH*, a collaborative scientific network between 12000 authors who submitted their articles to High Energy Physics category of arXiv e-print archive [46].
- *Twitter*, a maximum component of the Twitter social network consisting of more than 65000 users [47].

The descriptive statistics of the three additional networks are summarized in Table A1.

The results of numerical simulations using the networks listed in Table A1 are shown in Figs. A2–A4. After immunization, we examine the largest fraction of the network that can get infected (F) and the recovered fraction of nodes in an SIR epidemic (P_r). In all instances, the results are displayed as the functions of the fraction of immunized nodes (q). AS turns out to be a clear winner of these tests, outperforming its well-known competitors with a considerable margin. Overall, numerical tests strongly support the conclusion that AS offers new and improved power to identify nodes critical for the spread of infectious agents in network epidemiology.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (Grant no. 61174022), Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20131102130002), Research & Development Program of China (Grant no. 2012BAH07B01), National High Technology Research & Development Program of China (863 Program; Grant no. 2013AA013801), the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant no. BUAA-VR-14KF-02), and Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship Program for Foreign Researchers (no. P13380) and an accompanying Grant-in-Aid for Scientific Research.

[1] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200.

[2] A. Saumell-Mendiola, M. Á. Serrano, M. Boguñá, Epidemic spreading on interconnected networks, Phys. Rev. E 86 (2012) 026106.

[3] J. Zhou, G. Xiao, S. A. Cheong, X. Fu, L. Wong, S. Ma, T. H. Cheng, Epidemic reemergence in adaptive complex networks, Phys. Rev. E 85 (2012) 036107.

[4] Y. Wang, G. Xiao, Epidemics spreading in interconnected complex networks, Phys. Lett. A 376 (2012) 2689–2696.

[5] C. Buono, L. G. Alvarez-Zuzek, P. A. Macri, L. A. Braunstein, Epidemics in partially overlapped multiplex networks, PLOS One 9 (2014) e92200.

[6] J. Balthrop, S. Forrest, M. E. J. Newman, M. M. Williamson, Technological networks and the spread of computer viruses, Science 304 (2004) 527–529.

[7] C. T. Bauch, D. J. Earn, Vaccination and the theory of games, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 13391–13394.

[8] C. M. Schneider, T. Mihaljev, S. Havlin, H. J. Herrmann, Suppressing epidemics with a limited amount of immunization units, Phys. Rev. E 84 (2011) 061911.

[9] L. Hébert-Dufresne, A. Allard, J.-G. Young, L. J. Dubé, Global efficiency of local immunization on complex networks, Sci. Rep. 3 (2013) 2171.

[10] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, A. Vespignani, Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions, PLOS Med. 4 (2007) e13.

[11] N. Perra, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Activity driven modeling of time varying networks, Sci. Rep. 2 (2012) 469.

[12] R. Pastor-Satorras, A. Vespignani, Immunization of complex networks, Phys. Rev. E 65 (2002) 036104.

[13] B. Mirzasoleiman, M. Babaei, M. Jalili, Immunizing complex networks with limited budget, EPL (Europhys. Lett.) 98 (2012) 38004.

[14] C. M. Schneider, T. Mihaljev, H. J. Herrmann, Inverse targeting–an effective immunization strategy, EPL (Europhys. Lett.) 98 (2012) 46002.

[15] R. Cohen, S. Havlin, D. Ben-Avraham, Efficient immunization strategies for computer networks and populations, Phys. Rev. Lett. 91 (2003) 247901.

[16] L. K. Gallos, F. Liljeros, P. Argyrakis, A. Bunde, S. Havlin, Improving immunization strategies, Phys. Rev. E 75 (2007) 045104.

[17] X. Fu, M. Small, D. M. Walker, H. Zhang, Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization, Phys. Rev. E 77 (2008) 036113.

[18] L. Lü, Y.-C. Zhang, C. H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLOS One 6 (2011) e21202.

[19] K. Gong, M. Tang, P. M. Hui, H. F. Zhang, D. Younghae, Y.-C. Lai, An efficient immunization strategy for community networks, PLOS One 8 (2013) e83489.

[20] W. Wang, M. Tang, H.-F. Zhang, H. Gao, Y. Do, Z.-H. Liu, Epidemic spreading on complex networks with general degree and weight distributions, Phys. Rev. E 90 (2014) 042803.

[21] S. Yan, S. Tang, S. Pei, S. Jiang, Z. Zheng, Dynamical immunization strategy for seasonal epidemics, Phys. Rev. E 90 (2014) 022808.

[22] T. Nakagaki, H. Yamada, Á. Tóth, Intelligence: Maze-solving by an amoeboid organism, Nature 407 (2000) 470–470.

[23] A. Tero, R. Kobayashi, T. Nakagaki, A mathematical model for adaptive transport network in path finding by true slime mold, J. Theor. Biol. 244 (2007) 553–564.

[24] A. Tero, K. Yumiki, R. Kobayashi, T. Saigusa, T. Nakagaki, Flow-network adaptation in physarum amoebae, Theor. Biosci. 127 (2008) 89–94.

[25] A. Tero, T. Nakagaki, K. Toyabe, K. Yumiki, R. Kobayashi, A method inspired by physarum for solving the steiner problem, (Int. J. Unconv. Comput. 6 (2010) 109–123.

[26] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, T. Nakagaki, Rules for biologically inspired adaptive network design, Science 327 (2010) 439–442.

[27] S. Watanabe, A. Tero, A. Takamatsu, T. Nakagaki, Traffic optimization in railroad networks using an algorithm mimicking an amoeba-like organism, *Physarum plasmodium*, Biosystems 105 (2011) 225–232.

[28] L. Liu, Y. Song, H. Zhang, H. Ma, A. Vasilakos, Physarum optimization: A biology-inspired algorithm for the steiner tree problem in networks, IEEE Trans. Comput. 64 (2015) 819–832.

[29] Y. Song, L. Liu, H. Ma, A. Vasilakos, A biology-based algorithm to minimal exposure problem of wireless sensor networks, IEEE Trans. Netw. Serv. Manag. 11 (2014) 417–430.

[30] X. Zhang, Y. Zhang, D. Wei, Y. Deng, Solving shortest path problems with interval arcs based on an amoeboid organism algorithm, J. Inf. Comput. Sci. 9 (2012) 2081–2088.

[31] X. Zhang, S. Huang, Y. Hu, Y. Zhang, S. Mahadevan, Y. Deng, Solving 0-1 knapsack problems based on amoeboid organism algorithm, Appl. Math. and Comput. 219 (2013) 9959–9970.

[32] C. Gao, D. Wei, Y. Hu, S. Mahadevan, Y. Deng, A modified evidential methodology of identifying influential nodes in weighted networks, Physica A 392 (2013) 5490–5500.

[33] A. Adamatzky, Slime mould electronic oscillators, Microelectron. Eng. 124 (2014) 58–65.

[34] A. Adamatzky, T. Schubert, Slime mold microfluidic logical gates, Mater. Today 17 (2014) 86–91.

[35] T. Nakagaki, H. Yamada, A. Toth, Path finding by tube morphogenesis in an amoeboid organism, Biophys. Chem. 92 (2001) 47–52.

[36] S. Chae, S. K. Nguang, Sos based robust h-infinity fuzzy dynamic output feedback control of nonlinear networked control systems, IEEE Trans. Cybern. 44 (2014) 1204–1213.

[37] H. Zhang, C. Qin, B. Jiang, Y. Luo, Online adaptive policy learning algorithm for h-infinity state feedback control of unknown affine nonlinear discrete-time systems, IEEE Trans. Cybern. 44 (2014) 2706–2718.

[38] V. Bonifaci, K. Mehlhorn, G. Varma, Physarum can compute shortest paths, J. Theor. Biol. 309 (2012) 121–133.

[39] Z. Wang, A. Szolnoki, M. Perc, Interdependent network reciprocity in evolutionary games, Sci. Rep. 3 (2013) 1183.

[40] Z. Wang, A. Szolnoki, M. Perc, Optimal interdependence between networks for the evolution of cooperation, Sci. Rep. 3 (2013) 2470.

[41] L. Becchetti, V. Bonifaci, M. Dirnberger, A. Karrenbauer, K. Mehlhorn, Physarum can compute shortest paths: Convergence proofs and complexity bounds, in: Automata, Languages, and Programming, Springer, 2013, pp. 472–483.

[42] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, H. A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (2010) 888–893.

[43] P. Erdős, A. Rényi, On random graphs I, Publ. Math. Debrecen 6 (1959) 290–297.

[44] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (2002) 47.

[45] M. E. Newman, S. Forrest, J. Balthrop, Email networks and the spread of computer viruses, Phys. Rev. E 66 (2002) 035101.

[46] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Trans. Knowl. Discov. Data 1 (2007) 2.

[47] D. J. Watts, S. H. Strogatz, Collective dynamics of "small-world" networks, Nature 393 (1998) 440–442.

[48] C. Von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, P. Bork, Comparative assessment of large-scale data sets of protein–protein interactions, Nature 417 (2002) 399–403.

[49] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, Nature 406 (2000) 378–382.

[50] M. E. Newman, Spread of epidemic disease on networks, Phys. Rev. E 66 (2002) 016128.

[51] C. Castellano, R. Pastor-Satorras, Thresholds for epidemic spreading in networks, Phys. Rev. Lett. 105 (2010) 218701.

[52] C. Xia, L. Wang, S. Sun, J. Wang, An sir model with infection delay and propagation vector in complex networks, Nonlin. Dyn. 69 (2012) 927–934.

[53] Z. Wang, L. Wang, A. Szolnoki, M. Perc, Evolutionary games on multilayer networks: a colloquium, Eur. Phys. J. B 88 (2015) 1–15.

[54] W.-B. Du, X.-L. Zhou, M. Jusup, Z. Wang, Physics of transportation: Towards optimal capacity using the multilayer network framework, Sci. Rep. 6 (2016) 19059.

[55] M. Perc, A. Szolnoki, Coevolutionary games—a mini review, BioSystems 99 (2010) 109–125.

[56] Z. Wang, M. A. Andrews, Z.-X. Wu, L. Wang, C. T. Bauch, Coupled disease–behavior dynamics on complex networks: A review, Phys. Life Rev. 15 (2015) 1–29.

[57] B. Wei, J. Liu, D. Wei, C. Gao, Y. Deng, Weighted k-shell decomposition for complex networks based on potential edge weights, Physica A 420 (2015) 277–283.

[58] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, A. Arenas, Models of social networks based on social distance attachment, Phys. Rev. E 70 (2004) 056122.

[59] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E 72 (2005) 027104.

[60] Z. Lü, W. Huang, Iterated tabu search for identifying community structure in complex networks, Phys. Rev. E 80 (2) (2009) 026130.

Figures

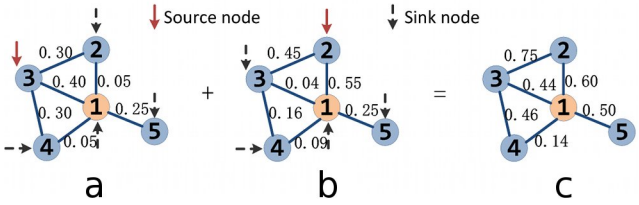

Figure 1: Internal functioning of AS. Node 1 marked with wax-yellow color is the focal node, whereas the other nodes are its first-level neighbors. Panels a and b show the results of two model iterations in which nodes 3 and 2 are selected as the source nodes, respectively. These results are summed to produce the so-called flow matrix in panel c. Upon running all possible iterations (see the accompanying text), the *AS*-score for the focal node is obtained by summing the elements of the flow matrix across all edges.

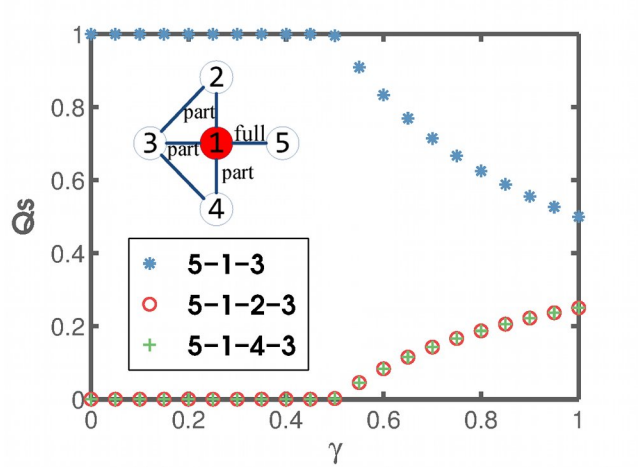

Figure 2: The role of the noise factor and irrational path selection. The distribution of flows, Qs, along the different paths from node 5 to node 3 (in the inset) is shown as a function of the noise factor, γ. If $\gamma > 0.5$, in addition to shortest path 5-1-3, some flow is equally distributed between the two remaining paths, 5-1-2-3 and 5-1-4-3, meaning that the relative importance of disease spreading pathways is correctly distinguished.

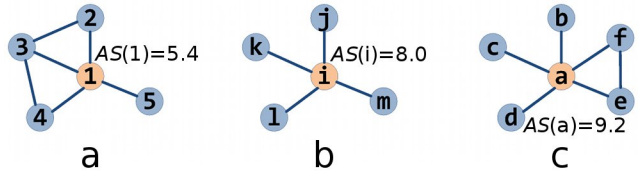

Figure 3: The advantage of AS over the other comparable strategies. Schematic network representations in which the node marked with wax-yellow color is the focal node, whereas the other nodes are its first-level neighbors. DCS fails to distinguish between the focal nodes in panels a and b. BCS fails to do the same for the focal nodes in panels b and c. AS, by contrast, does not suffer from the same problems as illustrated by displayed *AS*-scores.

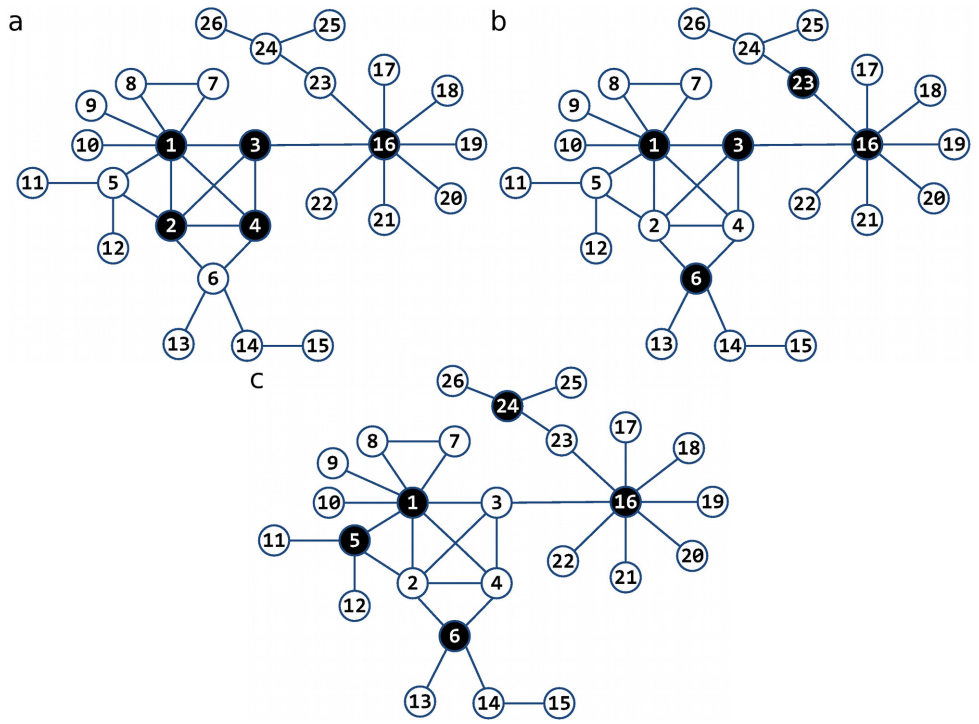

Figure 4: The performance of the considered immunization strategies in a simple, yet non-trivial network taken from Ref. [42]. Nodes marked for immunization by DCS and CCS (panel a), BCS (panel b), and AS (panel c). The best performer is AS in panel c because the biggest cluster left exposed after immunization is of size three (nodes 2, 3, and 4), whereas DCS and CCS in panel a leave the cluster of size four exposed (e.g. nodes 23 to 26) and BCS in panel b leaves the cluster of size 5 (nodes 2, 4, 5, 11, and 12) exposed to a disease.

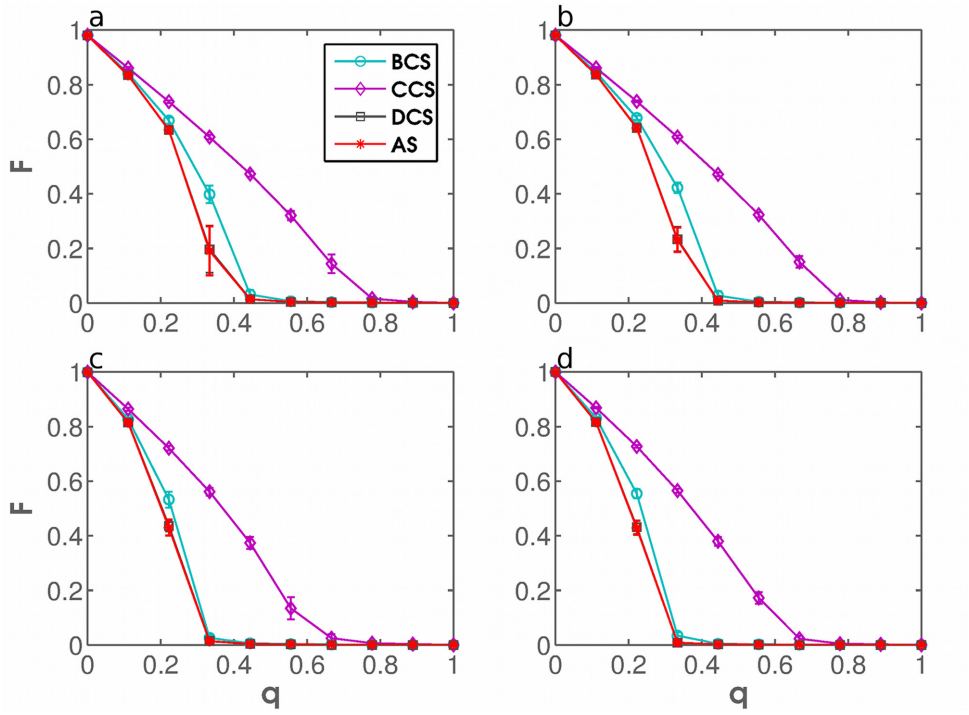

Figure 5: Performance testing in stylized, computer-generated networks. The fraction of the network that can be infected, F, is shown as a function of the fraction of immunized nodes, q, for DCS, BCS, CCS, and AS using ER network with $\langle k \rangle = 4.13$ (panel a), ER network with $\langle k \rangle = 4.03$ (panel b), SF network with $\alpha = 2.576$ and $\langle k \rangle = 6$ (panel c), and SF network with $\alpha = 2.468$ and $\langle k \rangle = 6$ (panel d). The error bars are derived from multiple realizations of the same simulation. The effectiveness of AS approximately matches that of DCS and outperforms that of BCS and CCS in random and scale-free networks likewise.

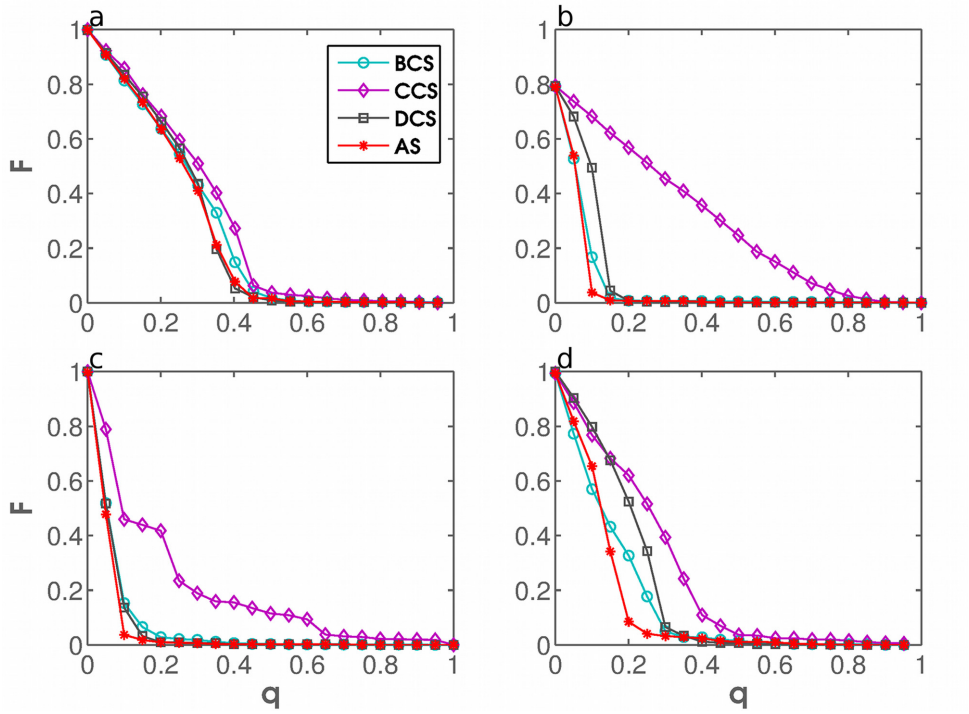

Figure 6: Performance testing in real-world, empirical networks. Fraction of the network that can be infected, F, is shown as a function of the fraction of immunized nodes, q, for DCS, BCS, CCS, and AS using the Email network (panel a), the GR-QC network (panel b), the Power network (panel c), and the Yeast network (panel d).

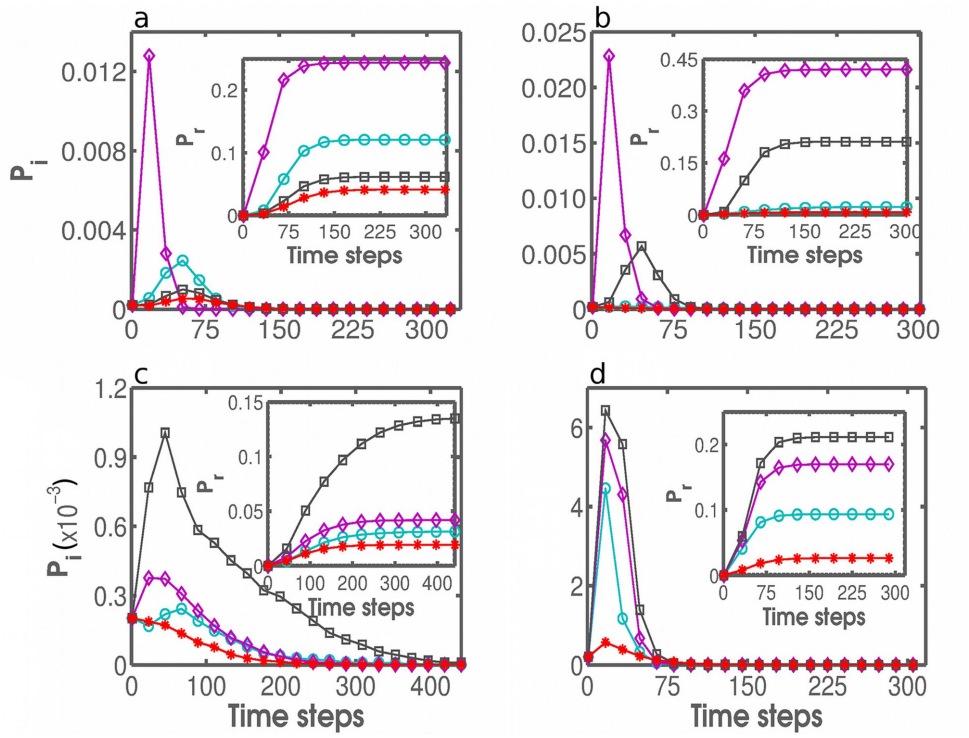

Figure 7: Performance testing in conjunction with SIR modeling. The infected fraction, P_i, and the recovered fraction, P_r, of nodes as the functions of time in SIR simulations. Shown are the comparisons between DCS, BCS, CCS, and AS (color coding as in Figs. 5 and 6) after having immunized fraction $q = 0.15$ of nodes in an SF network with $\langle k \rangle = 6$ (panel a), $q = 0.08$ of nodes in the GR-QC network (panel b), $q = 0.05$ of nodes in the Power network (panel c), and $q = 0.15$ of nodes in the Yeast network (panel d).

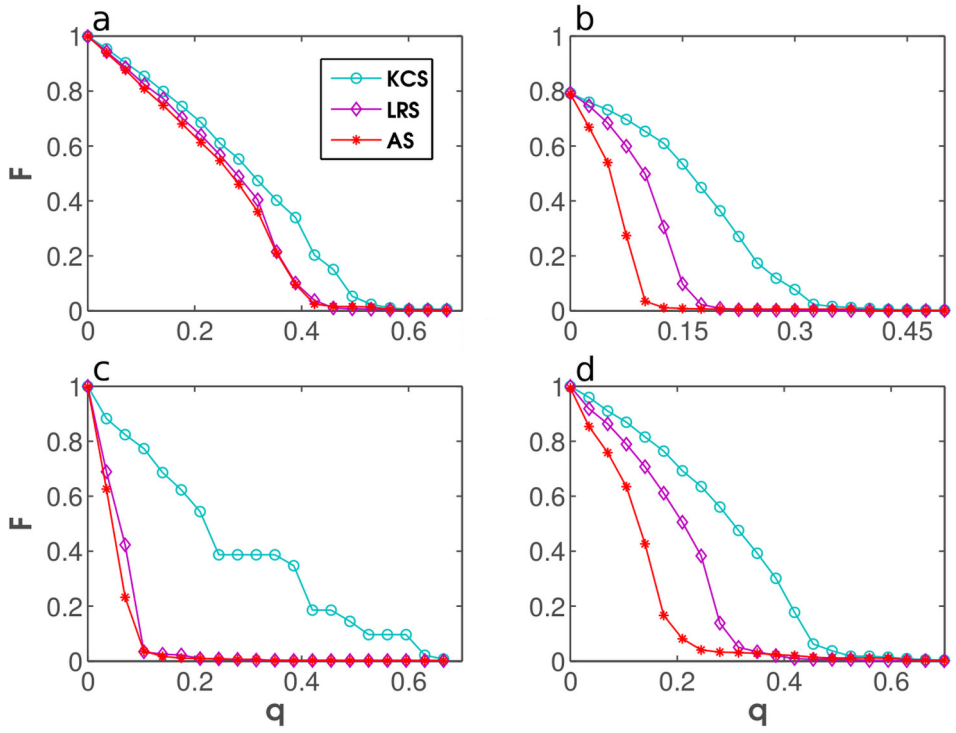

Figure A1: Comparison of AS with two recent immunization strategies. Fraction of the network that can be infected, F, is shown as a function of the fraction of immunized nodes, q, for KCS (k-core strategy), LRS (leader rank strategy), and AS using the Email network (panel a), the GR-QC network (panel b), the Power network (panel c), and the Yeast network (panel d).

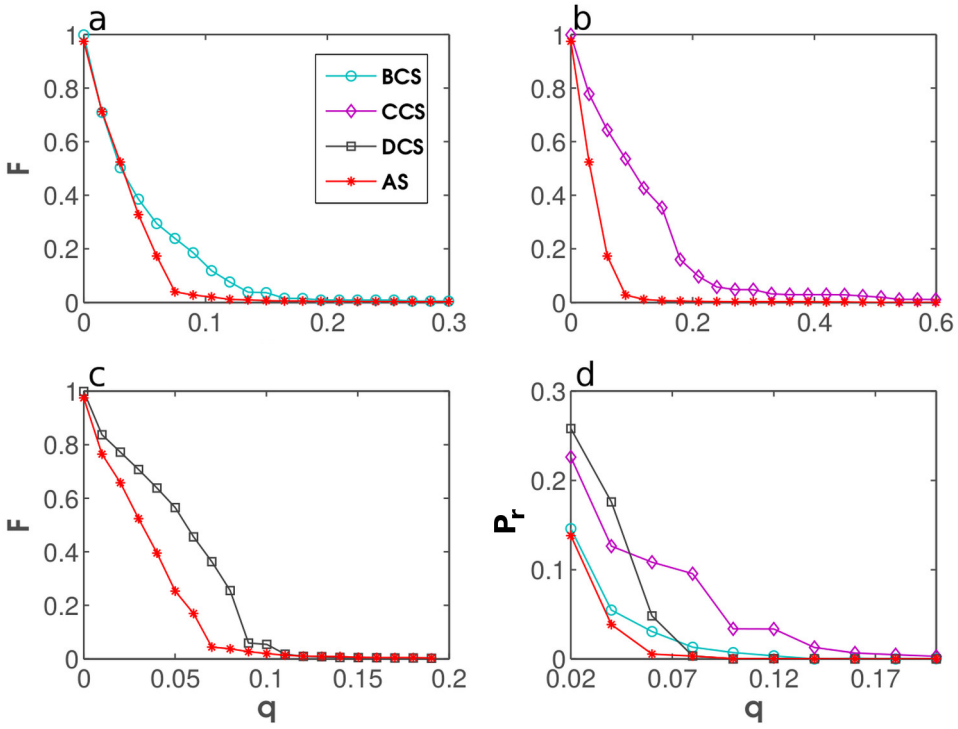

Figure A2: Performance testing in additional empirical networks: the case of the PGP network. Fraction of the network that can be infected, F (panels a–c), and the recovered fraction, P_r, of nodes in SIR simulations (panel d) are shown as the functions of the fraction of immunized nodes, q, for DCS, BCS, CCS, and AS.

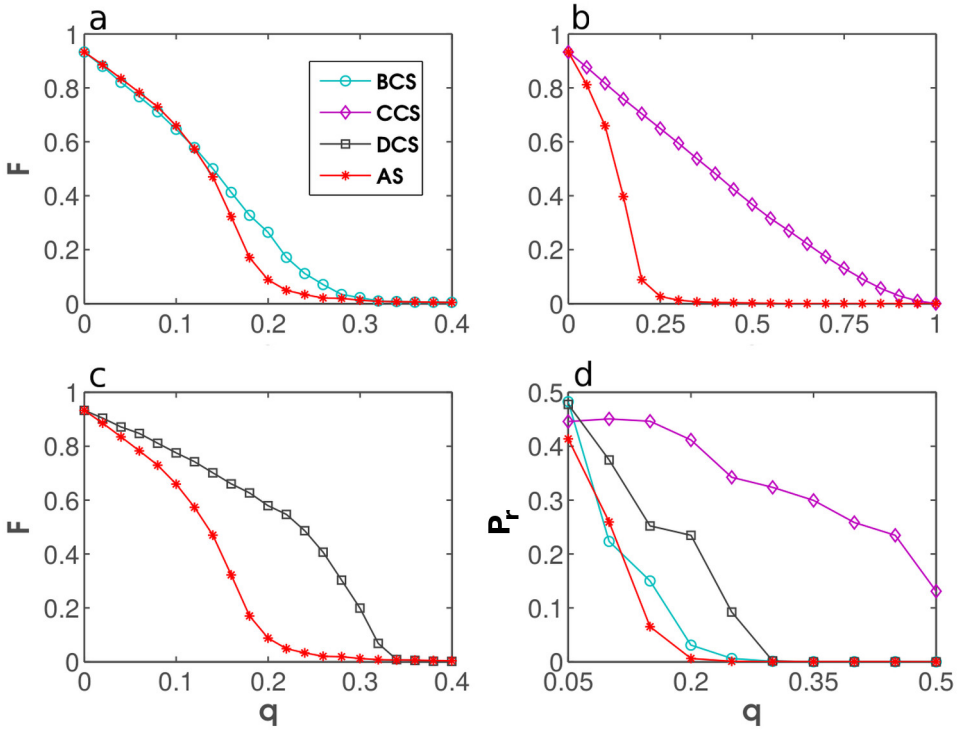

Figure A3: Performance testing in additional empirical networks: the case of the CA-HepPH network. Fraction of the network that can be infected, F (panels a–c), and the recovered fraction, P_r, of nodes in SIR simulations (panel d) are shown as the functions of the fraction of immunized nodes, q, for DCS, BCS, CCS, and AS.

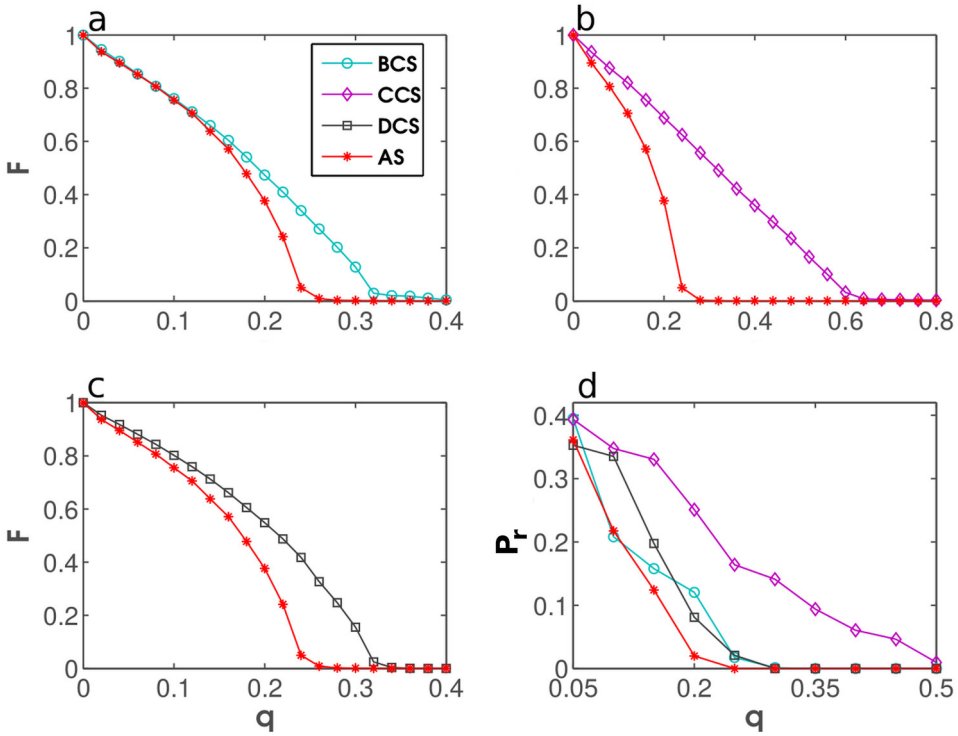

Figure A4: Performance testing in additional empirical networks: the case of the Twitter network. Fraction of the network that can be infected, F (panels a–c), and the recovered fraction, P_r, of nodes in SIR simulations (panel d) are shown as the functions of the fraction of immunized nodes, q, for DCS, BCS, CCS, and AS.

Tables

Table 1: Descriptive statistics of the four empirical networks used in performance testing.

Network	n	m	k_{max}	H	$\langle C \rangle$
Email	1133	5451	71	1.9421	0.2202
GR-QC	5242	14490	81	3.0512	0.5296
Power	4941	6594	19	1.4504	0.0801
Yeast	2375	11693	118	3.4756	0.3057

Quantities: network size (n), number of edges (m), maximum degree (k_{max}), degree heterogeneity ($H = \langle k^2 \rangle / \langle k \rangle^2$), and clustering coefficient ($\langle C \rangle$).

Table A1: Descriptive statistics of the three additional empirical networks used in performance testing.

Network	n	m	k_{max}	H	$\langle C \rangle$
PGP	10680	24316	205	4.1465	0.2659
CA-HepPH	12008	118505	491	6.5830	0.6115
Twitter	65911	185737	233	2.5829	0.1823

Quantities: network size (n), number of edges (m), maximum degree (k_{max}), degree heterogeneity ($H = \langle k^2 \rangle / \langle k \rangle^2$), and clustering coefficient ($\langle C \rangle$).